

## FORECASTING MEAN TEMPERATURE USING SARIMA MODEL FOR JUNAGADH CITY OF GUJARAT

D. K. DWIVEDI, G. R. SHARMA & S. S. WANDRE

Department of Soil and Water Engineering College of Agricultural Engineering and  
Technology Junagadh Agricultural University Junagadh, India

### ABSTRACT

*Forecasting of temperature can be done by combination of several mathematical models, using time series analysis. ARIMA (Auto Regressive Integrated Moving Average) model considers past data and prediction errors and relates its present data to obtain forecast. If time series has seasonality inherent in it, then Season ARIMA models are utilized to make forecast. Mean temperature forecasting for Junagadh city was carried out using SARIMA model, by using the past data from the period of 1984 to 2015. The orders of the model were estimated from the autocorrelogram and partial autocorrelogram. Several candidate models were developed for forecasting mean temperature. The model with the least value of Akaike Information Criterion (AIC) was selected as the appropriate model for forecasting mean temperature. SARIMA (1, 0, 1) (1, 1, 1) (12) was selected as the best model for mean temperature forecast. Test for normality of residuals were performed to check the adequacy of the selected model. According to the model diagnostics, the model was reliable for forecasting mean temperatures.*

**KEYWORDS:** Box Jenkins, Seasonal Auto Regressive Integrated Moving Average Model, Mean Temperature, Forecast, Junagadh

**Received:** May 27, 2017; **Accepted:** Jun 18, 2017; **Published:** Jul 11, 2017; **Paper Id.:** IJASRAUG201723

### INTRODUCTION

This study deals with the investigation of past mean temperature data recorded in Junagadh city, from 1984 to 2015. Seasonal Autoregressive Integrated Moving Average model was developed to obtain the forecast of 2016 monthly mean temperatures. A mathematical model representing stochastic process is called, time series model. Several methods can be used for forecasting mean monthly temperature, but stochastic modelling of temperature has proved to be a reliable method.

Kaushik and Singh (2008) predicted temperature and rainfall on a seasonal time scale of Mirzapur, Uttar Pradesh using 12 years data from 1994 to 2006. They used the Box-Jenkins time series seasonal ARIMA (Auto Regression Integrated Moving Average) approach for prediction of temperature and rainfall on monthly scales. The Box Jenkins technique was applied to predict temperature and rainfall of next five years by analysing last twelve years data (1994-2006). The results indicated that the seasonal ARIMA model provided reliable and satisfactory predictions for rainfall and temperature parameters on monthly scale.

Changet *al.* (2012) applied the seasonal ARIMA model to the time series precipitation data from 1961 to 2011, for forecasting monthly precipitation in Yantai, China. They found that the model SARIMA (1, 0, 1) (0, 1, 1)12 fitted the past data and could be used successfully for forecasting. AsamoahBoaheng (2014) forecasted temperature by using SARIMA model in Ashanti region by analysing the past data from 1980 to 2013, and

concluded that the best model for forecasting was SARIMA (2,1,1) (1,1,2)(12). Mills (2014) successfully modelled monthly temperatures of Kefalonia using multiplicative seasonal ARIMA model. Kibunja *et al.* (2014) forecasted precipitation of Mt. Kenya region by SARIMA model. The modelling was carried out based on Box Jenkins methodology, which involved model identification, parameter estimation and diagnostic check. Their model was found to be adequate for forecasting, as it also passed the normality tests of residuals. Dabralet *et al.* (2017) developed the “Seasonal Auto Regressive Integrative Moving Average” models (SARIMA) for monthly, weekly and daily monsoon rainfall time series and predicted monsoon time series for 14 years, from 2014 to 2027.

## MATERIALS AND METHODS

Temperature data recorded from 1984 to 2015 were obtained for Junagadh city, from the Agro meteorology department located in Junagadh Agricultural University, in Junagadh city of Gujarat. Monthly data in time series was analysed by SARIMA model for forecasting means temperature using XLSTAT.

SARIMA model requires the data to be stationary. Unit root test can be performed on the data to test whether the data is stationary or not. If it is not found to be stationary, then the data should be converted into stationary data. On shifting the time position to any arbitrary period, if the joint distribution does not change, then the data is said to be stationary data.

### Unit Root Test

Augmented Dickey fuller test or unit root test is used to determine whether the data series is stationary or non-stationary. The null hypothesis in this test states that the series has unit root and alternative hypothesis states that the series has no unit roots. The test statistic of the dickey fuller test is determined and the null hypothesis is rejected, if its value is found to be less than the critical value concluding that the data series has a unit root.

### Auto Regressive Integrated Moving Average Model

If the series is non-stationary, it is made stationary by differencing, and therefore, the term “integrated” may be used. Auto regressive terms in the forecasting equation refer to the lags of the series that has been differenced and moving average terms refer to the forecast error lags. An ARIMA model is usually written as ARIMA (p,d,q).

p=number of autoregressive terms

d=number of differences

q=number of moving average forecast error terms

Autoregressive model AR (p) can be written mathematically as

$$X_t = \alpha + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + \omega_t \quad (1)$$

Where,

$X_t$  = response variable at time t

$\alpha$  = constant term

$X_{t-1}, X_{t-2}, \dots, X_{t-p}$  = Response variable at time t-1, t-2 and t-p respectively

$\varphi_1, \varphi_2$  and  $\varphi_p$  = coefficients to be estimated

$\omega_t$  = error terms at time t

If the autocorrelogram dies down and the partial autocorrelogram cuts after lag q then AR (p) model should be used.

Moving average model MA (q) can be written mathematically as

$$X_t = \mu + \theta_1 \omega_{t-1} + \theta_2 \omega_{t-2} + \dots + \theta_q \omega_{t-q} + \omega_t \quad (2)$$

Where,

$X_t$  = response variable at time t

$\mu$  = constant mean of the process

$\omega_{t-1}, \omega_{t-2}, \dots, \omega_{t-q}$  = forecast errors at time series lags t-1, t-2 and t-q

$\theta_1, \theta_2$  and  $\theta_q$  = coefficients to be estimated

$\omega_t$  = error terms at time t

If the autocorrelogram cuts after lag p and the partial autocorrelogram dies down then MA (q) model should be used.

Autoregressive moving average model ARMA (p,q) can be written mathematically as

$$X_t = \alpha + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + \theta_1 \omega_{t-1} + \theta_2 \omega_{t-2} + \dots + \theta_q \omega_{t-q} + \omega_t \quad (3)$$

If the autocorrelogram cuts after lag p and the partial autocorrelogram dies down then MA (q) model should be used.

If both autocorrelogram and partial autocorrelogram dies down then ARMA (p, q) model should be used.

### Seasonal Auto Regressive Integrated Moving Average model

When both non-seasonal as well as seasonal factors are incorporated in a time series data then multiplicative seasonal ARIMA model may be used.

The SARIMA model can be expressed in following ways:

$$ARIMA(p, d, q) \times (P, D, Q) S \quad (4)$$

Where,

$p$  = non-seasonal Autoregressive order

$d$  = non-seasonal differencing

$q$  = non-seasonal Moving Average order

$P$  = seasonal Autoregressive order

$D$  = seasonal differencing

$Q$  = seasonal Moving Average order

$S$  = time span of repeating seasonal pattern

The non-seasonal components of AR and MA can be written mathematically as described in the ARIMA model, discussed previously. The seasonal components can be mathematically written as

Seasonal Autoregressive order:

$$\Phi(B^S) = 1 - \Phi_1 B^S - \dots - \Phi_p B^{pS} \quad (5)$$

Seasonal Moving Average order:

$$\Theta(B^S) = 1 + \Theta_1 B^S + \dots + \Theta_Q B^{QS} \quad (6)$$

Where,

$B^S$  = backshift operator at seasonal lag

$\Phi_1, \Phi_p$  = autoregressive coefficients to be determined at lag 1 and lag p

$\Theta_1, \Theta_Q$  = moving average coefficients to be determined at lag 1 and lag q

According to Box Jenkins methodology, model will be identified, its parameters will be estimated and then the model will be checked for its goodness of fit. If the model is adequate it may then be used for forecasting.

## RESULTS AND DISCUSSIONS

### Mean Temperature pattern (1984-2015)

Box Jenkins approach has been followed to analyse the rainfall data to build the forecast model. The mean temperatures recorded from January 1984 to December 2015 have been depicted in Figure 1. It is evident from the figure that the temperature series does not exhibit trend pattern, and so, it can be said to be stationary, however it exhibits periodic rise and fall.

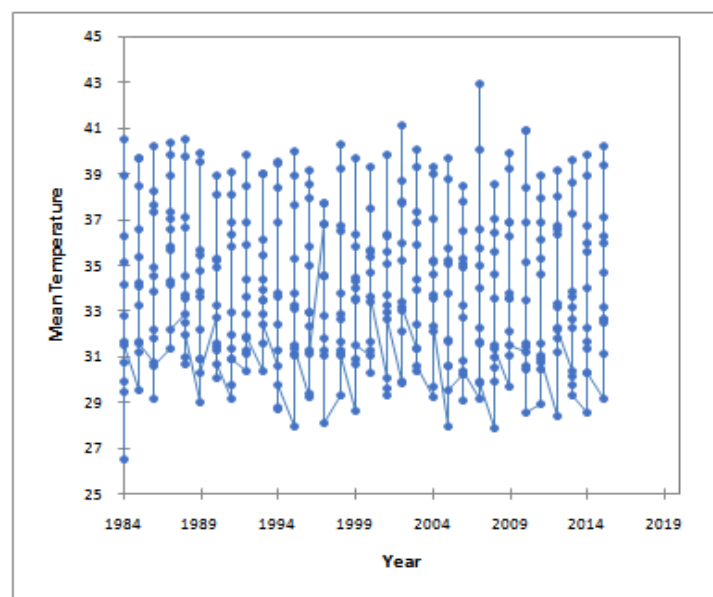


Figure 1: Time Series Plot of Mean Temperature Recorded From Jan. 1984-Dec. 2015

### Fitting an ARIMA Model of the Series Using the Box-Jenkins Approach

Box Jenkins methodology involves four steps

- Identification of model
- Estimation of model
- Checking goodness of fit of model
- Forecasting using the model

#### Identification of Model

Autocorrelation function plot and partial autocorrelation plots are analysed to check stationarity of data. In this case, the plotted ACF and PACF showed that the variance was constant, and therefore, they can be said to be stationary, however, spikes observed in the plots were out of the confidence limits of 95%. The plot of autocorrelogram showed that there was a seasonal pattern resembling a sinusoidal wave and the decay of spikes was gradual which meant that it had no trend and it was stationary. Augmented Dickey- Fuller (ADF) test was performed on the mean temperature series to check stationary which showed that the series was stationary (computed p-value was found to be less than the significance level of 5%). The series was differenced to eliminate the seasonality in series and the ACF and PACF were then analysed to determine the seasonal and non-seasonal orders of auto-regressive and moving average components of the model.

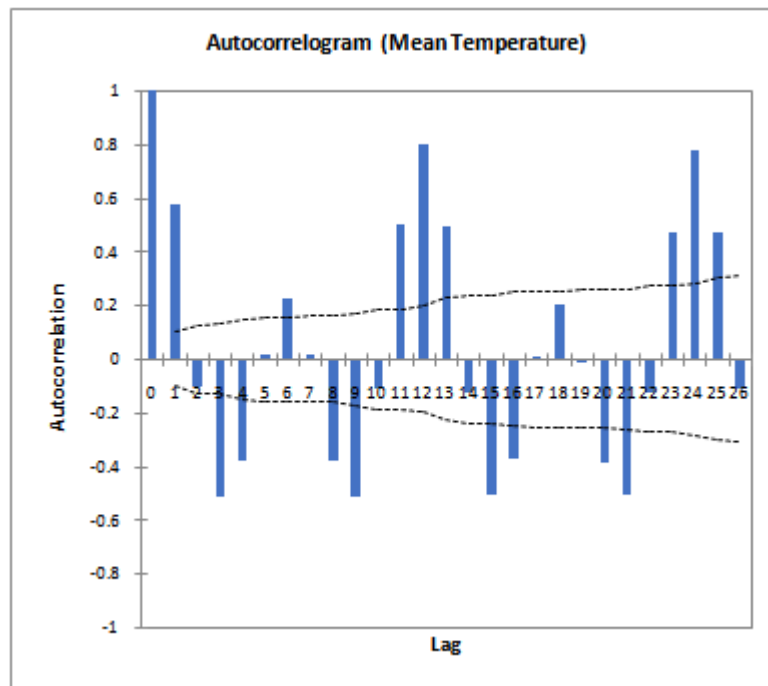


Figure 2: Autocorrelation Function (ACF) of the Series (Mean temperatures)

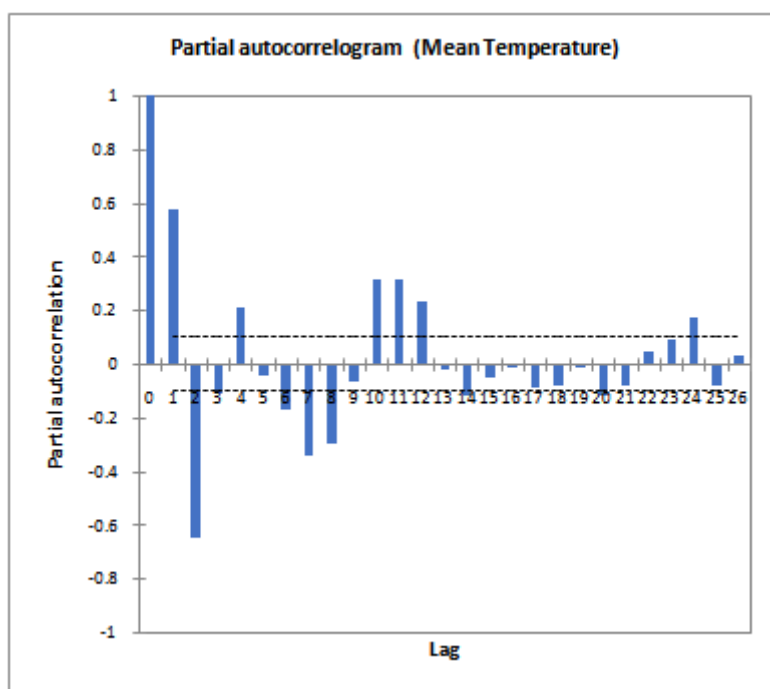


Figure 3: Partial Autocorrelation Function (ACF) of the Series (Mean temperatures)

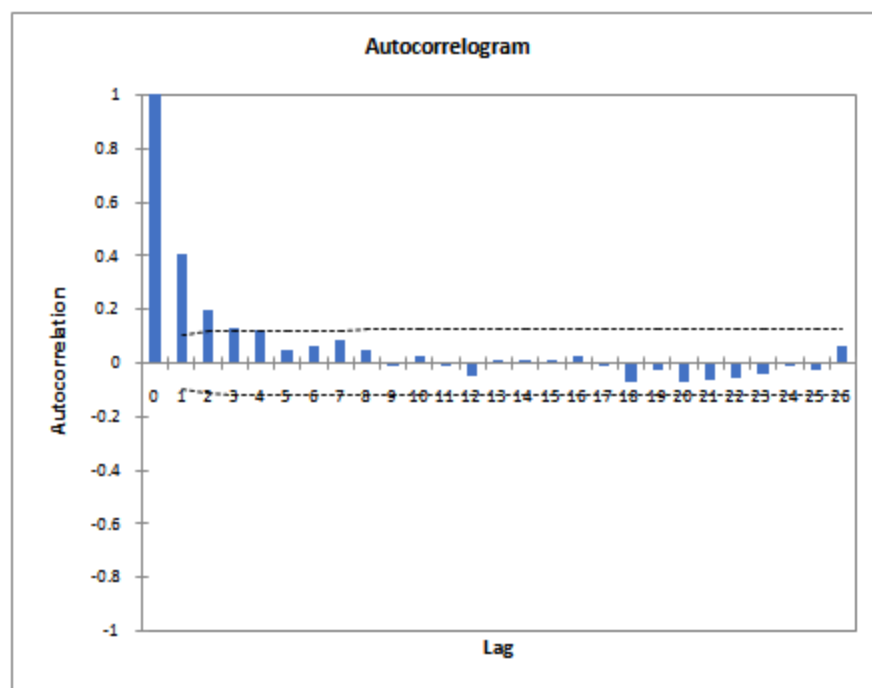
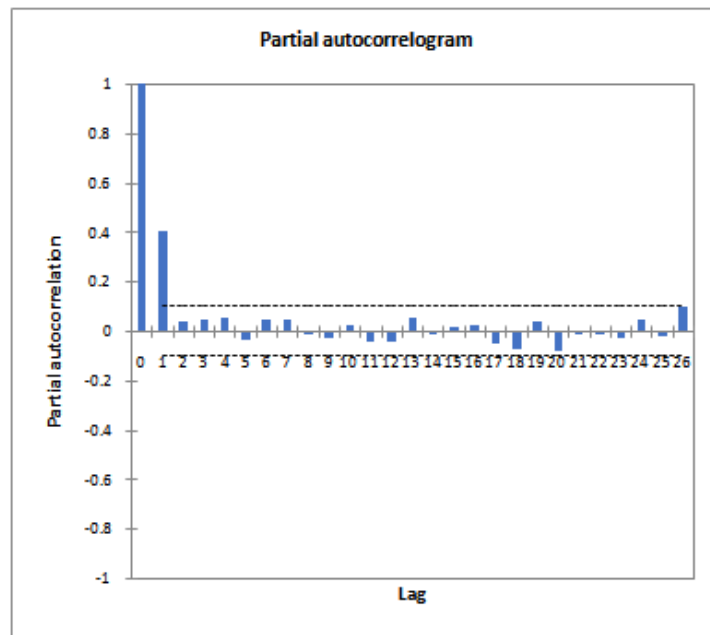


Figure 4: Autocorrelation Function (ACF) of One Seasonal Differencing



**Figure 5: Partial Autocorrelation Function (PACF) of One Seasonal Differencing**

The figure of ACF after one seasonal differencing shows that lags 1 and lags 2 are not zero as they are above the line of confidence limits, which means that the order of moving average (non-seasonal) could be up to 2. Spikes which are significant at lower lags showing that autoregressive terms (non-seasonal) may also be present.

#### Estimation of Model

Iterative procedure by using various orders were tested to obtain the least Akaike Information Criterion (AIC) value and the model with minimum value of AIC should be selected as the final model.

**Table 1: AIC for the Following trial SARIMA Models**

MODEL	AIC
SARIMA (1,0,2)(1,1,1)(12)	1256.146
SARIMA (1,1,2)(1,0,1)(12)	1857.026
SARIMA (1,0,1)(1,1,1)(12)	1255.531
SARIMA (2,0,1)(2,0,0)(12)	1882.548
SARIMA (1,0,1)(1,0,0)(12)	1881.082
SARIMA (2,0,0)(2,0,0)(12)	1902.463

**SARIMA (1,0,1)(1,1,1)(12)** has least AIC value indicating that it is the final selected model for predicting the mean temperature.

#### Goodness of Fit of Model

**Table 2: Parameter Estimation of the Final Model (SARIMA (1, 0, 1) (1, 1, 1)(12))**

Parameter	Value	Hessian Standard Error	Lower Bound (95%)	Upper Bound (95%)	Standard Error	Lower Bound (95%)	Upper Bound (95%)
AR(1)	0.538	0.123	0.296	0.779	0.053	0.434	0.642
SAR(1)	0.000	0.054	-0.105	0.105	0.063	-0.124	0.124
MA(1)	-0.149	0.147	-0.438	0.139	0.062	-0.272	-0.027
SMA(1)	0.000	0.143	-0.280	0.280	0.000	0.000	0.000

White noise (zero mean and constant variance) should be followed by the residuals. They can be checked by plotting the residuals as shown in following figures.

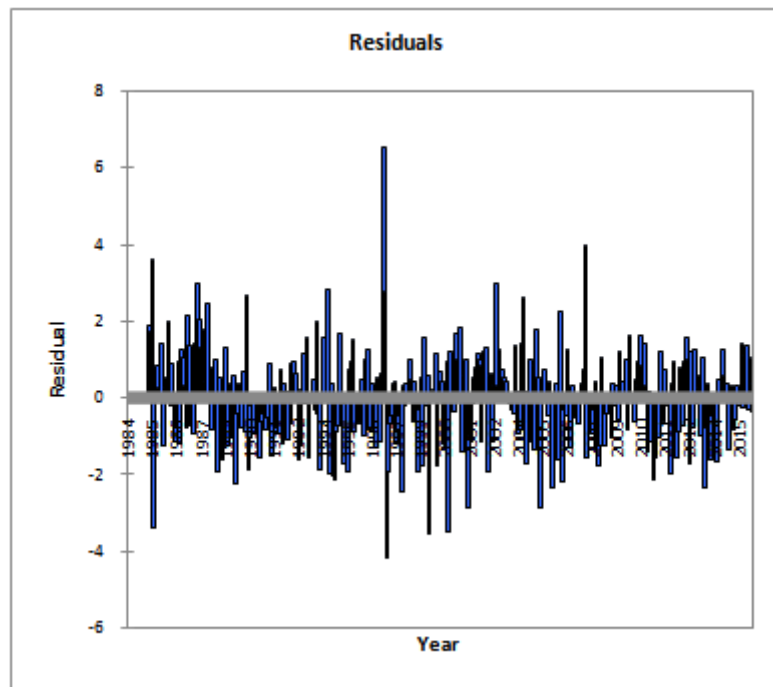


Figure 6: Plots of Model Residuals of Mean Temperatures

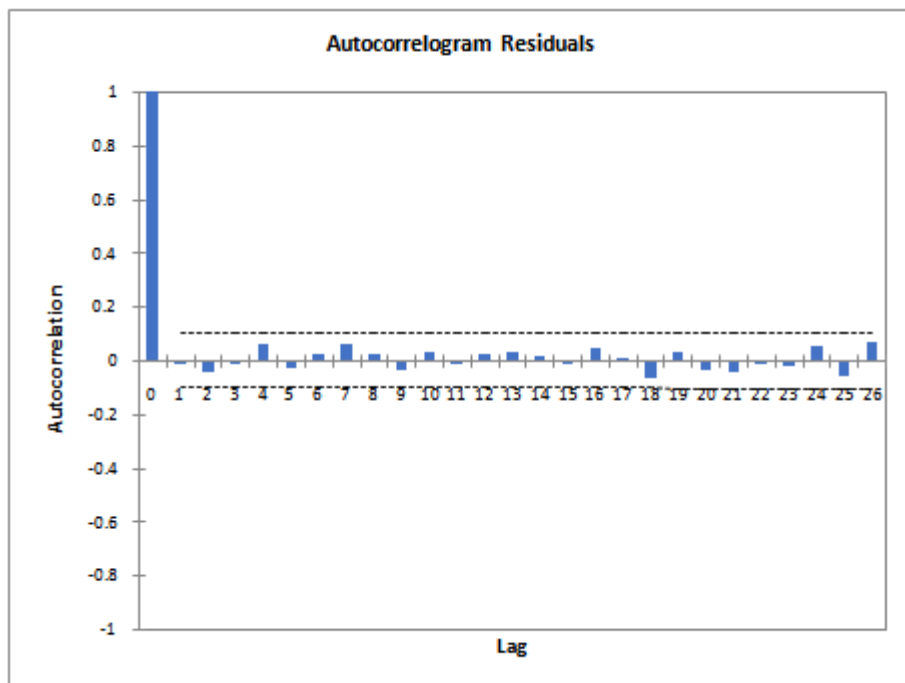
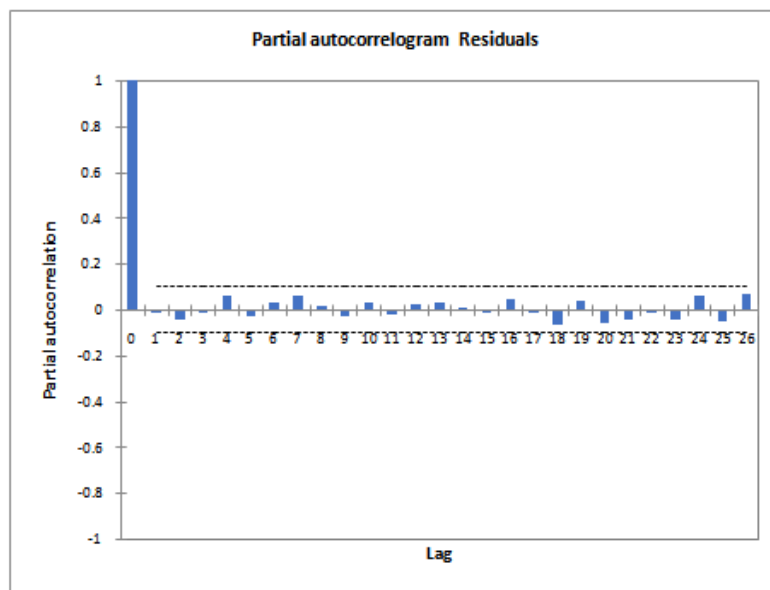


Figure 7: ACF of Residuals of Mean Temperature



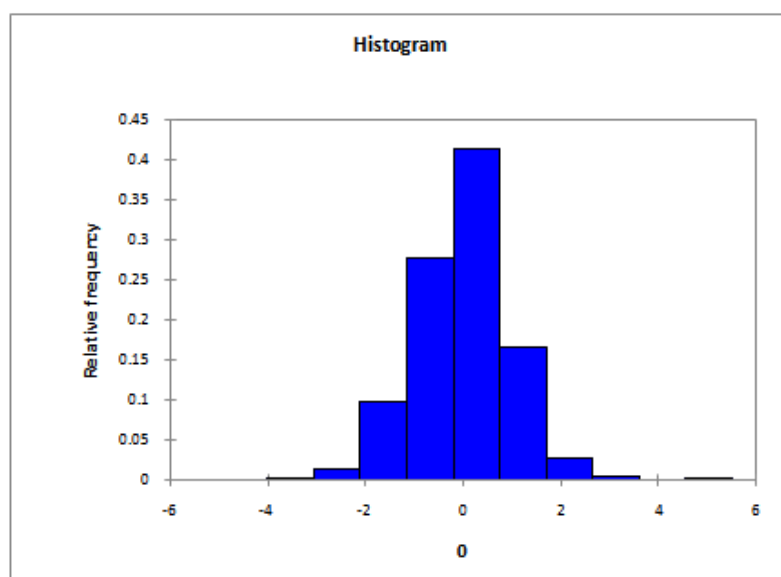


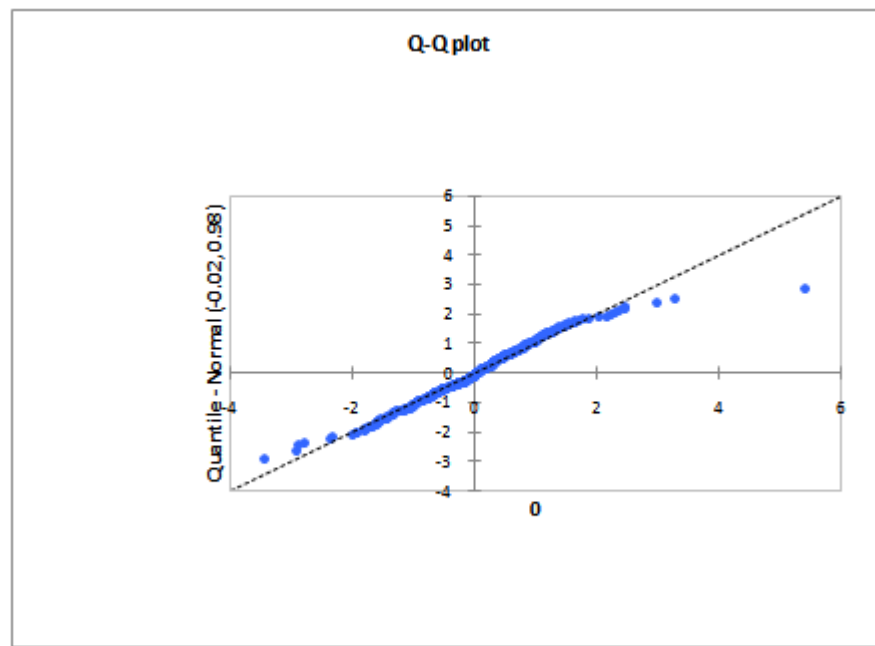
**Figure 8: PACF of Residuals of Mean Temperature**

From Figure 6 above, it is clear that the residuals follow white noise. Also, as shown in figure 7 and figure 8 the ACF and PACF of the residuals of the model shows that the autocorrelation of the residuals are all zero starting from lag 1 which means that they are uncorrelated. The required assumption of white noise is being followed by the final model SARIMA (1, 0, 1) (1, 1, 1) (12).

#### Normality Test for Residuals

Figure 9 shows the test plot for the test of normality of the residuals. It mostly follows a straight line, as observed in the Q-Q plot suggesting that the model residuals are normally distributed, and from histogram, it can be seen that it follows bell curve distribution (normal distribution). Hence, the SARIMA model SARIMA (1,0,1)(1,1,1)(12) follows the required assumption indicating that model can be used for forecasting.





**Figure 9: Plots of Normality Test for Residuals**

#### **SARIMA (1,0,1)(1,1,1) (12) Forecast**

The derived model was used for forecasting and the results are as shown in the following table.

**Table 5: Forecast Values for 12 Months Mean Temperatures for 2016**

Year	Month	Forecast (°C)	Confidence Limits	
			Lower	Upper
2016	January	30.263	27.705	32.821
	February	32.763	30.157	35.369
	March	36.621	34.001	39.241
	April	39.342	36.718	41.965
	May	38.849	36.225	41.474
	June	35.745	33.120	38.370
	July	31.418	28.793	34.044
	August	30.646	28.020	33.271
	September	32.638	30.013	35.263
	October	35.709	33.084	38.334
	November	33.807	31.182	36.433
	December	31.043	28.416	33.671

#### **CONCLUSIONS**

The set of data of mean temperatures of Junagadh from 1984 to 2015 was observed to be stationary, which was confirmed by the autocorrelation, partial autocorrelation plots and by conducting dickey fuller test. One seasonal differencing had to be carried out for eliminating the seasonality component in the time series of mean temperatures. Candidate models were developed as mentioned in model building process following Box Jenkins approach, and AIC values were obtained for each candidate models. The final model selected was SARIMA (1,0,1)(1,1,1)(12). Residuals were checked and they were found to be following white noise, which meant that they were uncorrelated. The residuals also passed the normality test, as could be seen from the histogram and Q-Q plot. Diagnosis of the model was performed, which

indicated that the model could be used for forecasting.

#### REFERENCES

1. Asamoah-Boaheng, M. (2014). Using SARIMA to Forecast Monthly Mean Surface Air Temperature in the Ashanti Region of Ghana. *International Journal of Statistics and Applications*, 4(6), 292-298.
2. Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.
3. Chang, X., Gao, M., Wang, Y., & Hou, X. (2012). Seasonal autoregressive integrated moving average model for precipitation time series. *Journal of Mathematics & Statistics*, 8(4).
4. Dabral, P. P., & Murry, M. Z. (2017). Modelling and Forecasting of Rainfall Time Series Using SARIMA. *Environmental Processes*, 1-21.
5. Kaushik, I., & Singh, S. M. (2008). Seasonal ARIMA model for forecasting of monthly rainfall and temperature. *Journal of Environmental Research and Development*, 3(2).
6. Kibunja, H. W., Kihoro, J. M., Orwa, G. O., & Yodah, W. O. (2014). Forecasting Precipitation Using SARIMA Model: A Case Study of Mt. Kenya Region.
7. Mills, T. C. (2014). Time series modelling of temperatures: an example from Kefalonia. *Meteorological Applications*, 21(3), 578-584.

